Revision Memo

An Election Forecasting Model for Subnational Elections

(Electoral Studies, JELS-D-24-00368)

March 14, 2025

Dear Editor,

We are thankful for the opportunity to revise and resubmit our manuscript *"An Election Forecasting Model for Subnational Elections"* to *Electoral Studies*. We thank you and the reviewers for the helpful and constructive comments, and we have carefully and thoroughly revised the manuscript.[1]

At the outset, we would like to outline how our revised manuscript addresses the main points highlighted by the editor. We have focused on clarifying our modelling approach, benchmarking the model's performance, and extending the literature review. We also evaluate additional models including the economic performance at the time of elections which might affect support especially for government parties.

We are confident that the manuscript has improved significantly after this round of revisions and that the overall changes have significantly strengthened the manuscript.

Sincerely

The Authors

[1] Reviewer comments have a running number and are written in italics, responses follow in normal font. The modifications we have made to the manuscript are marked in red in the memo and the revised document.

Response to Reviewers (without Author Details)

Response to Reviewer 1

Comment 1 *I recommend a Revise and Resubmit. The paper advances election forecasting studies at the subnational level, namely the state level. And the case study itself, German states, has value. However, the literature thins out at points, only 3 German states are targeted, its claims for Bayesian newness questionable, and the writing seems too much a technical report. Finally, the paper lacks pedagogy, being unnecessarily opaque. I elaborate on these points below, in page-by-page order.*

We thank Reviewer 1 for the suggestion to resubmit a revised version of our manuscript. The points considered really helped a lot to improve the clarity and structure of our paper.

Comment 2 *The authors taut "a new Bayesian election model....rather than simply aggregating polls." But that presents a restricted choice of approaches. I agree that we must do more than "simply aggregating polls." Indeed, we can build models that are theoretically driven by political science, and estimated by frequentist methods, such a classic linear regression. Nowhere in the paper do the authors make clear how the model set up would be superior, theoretically or empirically, to the more straightforward discussion in the political economy modelling literature of election forecasting, which is plentiful and represents common coin, whereby readers (and students) can handily follow the research story as it upfolds. Not so here, where we are often met with esoteric or "sophisticated" terms that strain accessibility, to no good purpose. Tellingly, the authors repeat themselves a good deal, without illuminating the practice.*

We appreciate the reviewer's comments and the opportunity to clarify our approach.

Our model follows a synthetic forecasting framework, integrating both structural and polling-based predictors (Lewis-Beck et al. 2016; Lewis-Beck & Dassonneville 2015a). Classic election forecasting models in political science emphasize structural factors—such as economic performance, incumbency, and approval ratings—as key predictors of electoral outcomes. In a multi-party context like Germany, the set of structural predictors differs slightly from those used in two-party context like U.S. elections. Since our goal is to present a general model applicable across various contexts, we do not predefine a fixed set of predictors. Instead, we allow for flexibility in their selection based on the specific electoral setting when introducing the model. For our application to German state elections, we than choose predictors that have been widely discussed in the German forecasting literature.

To clarify this approach, we have expanded the discussion in two parts of the manuscript: first, when introducing the general model, and second, when describing its application. In the theory section, we added the following sentences:

> **A Forecasting Model for Subnational Elections**
>
> In this section, we develop a forecasting model for subnational elections. Our model follows a synthetic forecasting framework, integrating both fundamental and polling-based predictors (Lewis-Beck et al. 2016; Lewis-Beck & Dassonneville 2015a).
>
> The target for our forecasting model are subnational election results. We assume that each observation v_i, where $i = (k, e) \in \mathcal{K} \times \mathcal{E}$, represents the vote share of a party $k \in \mathcal{K}$ in a subnational election $e \in \mathcal{E}$.[2] In applications, we select a number of relevant parties $k_j \in \mathcal{K}$ for each election e and subsume all other parties into a residual party $k_{res} \in \mathcal{K}$ called 'Others'.
>
> To build the polls-based part of our forecasting model, we have data from pre-election polls $p_{i,t}$, i.e. we have the share of voters in a poll published *before* election e that intend to vote for each party k. Since these vote shares potentially vary across each day $t \in \{1, \ldots, T_e\}$ between the previous and the upcoming election e, we collect all poll data for party k before a subnational election e in a row vector $\boldsymbol{p}_i = (p_{i,1}, \ldots, p_{i,T_e})$ of length T_e. Each entry $p_{i,t}$ represents the published poll-based vote shares for party k in election e if a poll was published on day t, or is set to missing otherwise. We further define the lead time at time t to election e as $l = T_e - t$ with $l \in \{1, \ldots, T_e\}$.
>
> (...)
>
> This framework allows us to estimate the latent support for each party before a subnational election over time, accounting for both the random evolution of actual support and the inherent noise in poll data. For

[2]The vote shares fall into the unit interval $0 < v_i < 1$. Additionally, the vote shares of all parties in any given election sum to one, i.e., $\sum v_i = 1$.

estimation, we only use data up to a specific lead time l to election e, such that we only consider polls that were published up to l days prior to that election. This subsets the data for each party to $\boldsymbol{p}_i = (p_{i,1}, \dots, p_{i,T_e-l})$, a vector of length $T_e - l$. Given a specific lead time, we estimate the latent support using a Kalman filter (West & Harrison 1997, p.103-107).[3] The advantage of using a dynamic linear model is that it provides latent support estimates for all lead times before an election, even if no polls were published on that day or within a nearby time-frame. This allows us to relate the latent support for a party to the election result at different lead times.[4]

Next, we integrate the poll-based with the fundamentals-based model to forecast the election results. We assume that the vote share of a party v_i is related to a set of observed fundamental predictors, and the latent support from the dynamic linear model with a specific lead time. The fundamental predictor variables are factors theoretically related to the support of political parties in upcoming elections, such as economic indicators, incumbency status, and government approval ratings.[5] We collect these observed fundamentals predictors in a matrix $\boldsymbol{X}$ that has N rows and the C predictors in respective columns, $\boldsymbol{x}_i$ is the row vector that hold the values of the predictors of v_i, the vote share of party k in an election e. In order to identify a constant term for the systematic component of the model, we add a column with ones to the C predictors in matrix $\boldsymbol{X}$. The latent support $\pi_{i,l}$ for party k before election e with a specific lead time l is taken from the dynamic linear model that is estimated based on data available at this lead time. We collect the support for all parties before election e with a given lead time in a column vector $\boldsymbol{\pi}_l$.

To ensure that our vote share forecasts remain within the 0% to 100% range, we apply a transformation to the dependent variable. We use a

[3] We estimate the model variances (observational error variance and evolution variance) using maximum likelihood routine and set uninformative priors on the initial latent states ($m_0 = 0$, $S_0 = 5$). The estimation is implemented using the R-package dlm (Petris 2010).

[4] If there are no polls available for a particular party at a specific lead time before the election, we cannot estimate the dynamic linear model, and the latent support for this party is marked as missing

[5] It is important to note that we define a general framework rather than pre-specifying a fixed set of fundamental predictors. The relevant predictors will vary depending on the specific application. For our application to German state elections, for example, we select predictors with a strong theoretical foundation in political science debates on voting behavior in Germany.

log ratio transformation for the observed election outcomes $\hat{v}_i = \ln \frac{v_i}{1-v_i}$ to ensure that estimated confidence intervals for the untransformed election outcomes fall within the unit interval. The linear regression model with log ratio transformed vote shares is defined as:

$$
\begin{aligned}
\hat{v}_i &\sim N(\mu_i, \sigma) & (1)\\
\mu_i &= \boldsymbol{x_i}\boldsymbol{\beta'} + \gamma\pi_{i,l} & (2)
\end{aligned}
$$

where $\boldsymbol{\beta}$ are the effects of the fundamental predictor variables, γ the effect of $\pi_{i,l}$, the latent party support with an election-specific lead time l, and the constant error variance σ. The effect parameters indicate how the expected log-ratio vote shares changes with a change in the fundamental predictor variables or the latent support in the polls. We collect the parameters of the model in a vector $\boldsymbol{\theta} = [\boldsymbol{\beta}, \gamma, \sigma]$.

We estimate the model using Bayesian methods.[6] The posterior distribution of the model parameters is proportional to the likelihood times the priors, while $\boldsymbol{X}$ and $\boldsymbol{\pi}_l$ is fixed in the likelihood and $\boldsymbol{v}$ is the vector of vote shares.

$$
P(\boldsymbol{\theta} \mid \boldsymbol{v}, \boldsymbol{X}, \boldsymbol{\pi}_l) \propto P(\boldsymbol{v} \mid \boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{\pi}_l)\, P(\boldsymbol{\theta}) \qquad (3)
$$

Bayesian estimation requires the specification of priors beliefs about the parameters of the model. The priors are defined in terms of a probability distribution $P(\boldsymbol{\theta})$. We generally assume pairwise independent distributions for$P(\boldsymbol{\theta}) \propto P(\boldsymbol{\beta})\, P(\gamma)\, P(\sigma)$ and use application-specific priors.

Based on the model, we can obtain a forecast for the upcoming election. We define the predicted vote shares for the relevant parties in the upcoming election as $\boldsymbol{v}^\star$ (excluding party k_{res})[7], $\boldsymbol{X}^\star$ holds the values of the fundamental predictor variables, and $\boldsymbol{\pi}_l^\star$ represents the estimated

[6] We use Bayesian estimation for its probabilistic interpretation, which allows direct probability statements about events (e.g., 5/6 probability that the results of a party will fall within the credible interval). This improves interpretability and facilitates communication, as shown in recent Bayesian election forecasts (Stoetzer et al. 2019; Kang & Oh 2024; Chen et al. 2023).

[7] For the forecasts, we leave out the support for parties in the residual category k_{res}. This implies that the remaining vote share for these other parties is set to the rest when the predicted vote shares of the relevant parties is subtracted from 100%.

latent support of the parties given a specific lead time. With this, we can compute the posterior predictive distribution.

$$P\left(\boldsymbol{v}^{\star} \mid \boldsymbol{X}^{\star}, \boldsymbol{\pi}_{l}^{\star}\right)=\int_{\boldsymbol{\theta}} P\left(\boldsymbol{v}^{\star} \mid \boldsymbol{X}^{\star}, \boldsymbol{\pi}_{l}^{\star}, \boldsymbol{\theta}\right) P\left(\boldsymbol{\theta} \mid \boldsymbol{v}, \boldsymbol{X}, \boldsymbol{\pi}_{l}\right) d\boldsymbol{\theta}. \quad (4)$$

The posterior predictive distribution represents the probability distribution of future vote shares given new predictor values, incorporating the uncertainty in our parameter estimates. It is obtained by integrating over the posterior distribution of the parameters, effectively averaging predictions across all plausible parameter values inferred from the observed data.

The posterior predictive distribution allows us to generate forecasts from the model. We can derive point estimates for election results using the posterior mean and construct credible intervals to quantify the inherent uncertainty in our predictions.

To implement the estimation and forecasting, we rely on Markov Chain Monte Carlo (MCMC) methods. We sample from the posterior distribution and the posterior predictive distribution to obtain forecasts for upcoming elections using the No-U-Turn sampler (NUTS) (Carpenter et al. 2017) as implemented in Stan, which we access using the R-package rstanarm (Goodrich et al. 2020).

We now also make it clear that we select a set of specific fundamentals predictors for the application at hand in the application section.

For applying the general model, it is crucial to select fundamental predictors relevant to the specific context. In our case, we choose predictor variables commonly used in forecasting models for German federal elections. First, government participation is a critical predictor, as incumbency generally confers electoral advantages. Voters often prefer incumbents, based on their perceived competence or continuity in governance (Allers et al. 2022; Eggers & Spirling 2017).

(...)

Comment 3 *The preregistration of ex ante forecasts is good. They go on to say "results fell within credible intervals..." Do they appear better than usual confidence intervals?*

The question of the usefulness of credible versus confidence intervals highlights another advantage of our forecasting approach. Bayesian estimation allows us to use credible intervals, which provide a direct probabilistic interpretation—an advantage over confidence intervals. A key benefit of the Bayesian framework is its ability to quantify uncertainty in election forecasts more intuitively than frequentist methods. Interpreting predictive intervals from a ordinary least square (OLS) regression in the frequentist framework requires an asymptotic interpretation—over infinitely repeated elections, a certain percentage of the generated confidence intervals will contain the election outcome—which is arguably cumbersome. In contrast, credible intervals allow for a straightforward probability interpretation: there is a specific probability that the result falls within the interval. This makes them particularly valuable in forecasting. For clarity, we elaborate on our choice of Bayesian methods in a footnote.

> We employ Bayesian estimation for its intuitive probabilistic interpretation of credibility intervals, which allows direct probability statements about events—unlike conventional confidence intervals. This approach is particularly effective in communicating uncertainty to a general audience. For instance, we report 5/6 credible intervals, meaning that the vote share of each party has a 5-in-6 probability of falling within this range—akin to rolling any number except a six on a fair die. This enhances interpretability and facilitates communication, as demonstrated in recent Bayesian election forecasts (Stoetzer et al. 2019; Kang & Oh 2024; Chen et al. 2023).

Comment 4 *The authors discuss the categories of structural models but the examples leave out much of the literature, a literature which, even at the state level, has existed for 20 years, e.g., see Kedron Bardwell's 2004 piece in PS, on "State-Level Forecasts of US Senate Elections."*

We appreciate the reviewer's suggestion and have expanded our discussion of structural models in subnational election forecasting to better reflect the breadth of existing literature. In particular, we have now included references to Bardwell & Lewis-Beck (2004), who developed a state-level model for forecasting U.S. Senate elections, as well as Linzer (2013) and Klarner (2010, 2018), who provide forecasts for U.S. presidential and state legislative elections, respectively. Additionally, we now discuss Hummel & Rothschild (2014), who model gubernatorial election outcomes based on state-level fundamentals. These additions add a more comprehensive representation of structural forecasting models at the subnational level.

> Several studies have developed structural models for forecasting subnational elections in the U.S. For example, Bardwell & Lewis-Beck (2004) presents a state-level model for forecasting U.S. Senate elections, while Linzer (2013) applies a Bayesian approach to predict U.S. presidential election outcomes at the state level. Similarly, Klarner (2010, 2018) introduce models for forecasting U.S. state legislative elections in 2010 and 2018, respectively. In the context of gubernatorial races, Hummel & Rothschild (2014) develops a model accounting for state-level fundamentals.

Comment 5 *The latent state equation and its development would benefit from more verbal elaboration. Further, it is repetitive, as is most of the text.*

We appreciate the reviewer's suggestion and have expanded the explanation of the latent state equation to improve clarity. Specifically, we now provide a more detailed discussion of the two equations and their role in capturing the underlying electoral support and how it evolves over time.

> We first devise a dynamic model to estimate the latent support for party k in a subnational unit prior to an election e, based on polling data. To do so, we employ a *dynamic linear model* with a random walk component (West & Harrison 1997).[8] The dynamic model consists of two key components: a *measurement equation*, which links observed polling data to the unobserved latent support, and a *latent state equation*, which describes how latent support evolves over time.
>
> The *measurement equation* is specified as follows:
>
> $$p_{i,t} = \pi_{i,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim \mathcal{N}(0, R_i) \tag{5}$$
>
> This equation states that the observed poll share $p_{i,t}$ for party k at time t consists of the latent true support $\pi_{i,t}$ plus an observation error

[8] Most poll dynamic models rely on dynamic linear models (Walther 2015) or apply transformations to adapt poll data to continuous measurement error assumptions (Stoetzer et al. 2019). For an alternative approach using non-linear state space models for polling data, see (Stoetzer & Orlowski 2020).

term $\epsilon_{i,t}$, which is assumed to be normally distributed with variance R_i. The term R_i reflects the uncertainty in individual polls, accounting for sampling variability and other sources of measurement error.

The *latent state equation*, which governs the temporal evolution of latent party support, is defined as:

$$\pi_{i,t} = \pi_{i,t-1} + \eta_{i,t}, \quad \eta_{i,t} \sim \mathcal{N}(0, Q_i) \tag{6}$$

Here, the latent support $\pi_{i,t}$ follows a *random walk*, meaning that the best predictor for current support is simply the previous period's support plus a stochastic evolution term $\eta_{i,t}$. This formulation assumes that changes in support occur incrementally over time rather than experiencing sudden jumps, making it well-suited for modeling gradual shifts in voter support. The variance Q_i captures the degree of expected change in latent support over time and is specific to each party-election combination. To initialize the process, we assume an initial latent state: $\pi_{i,0} \sim \mathcal{N}(m_0, S_0)$, where m_0 represents the prior expectation of support before polling data is observed, and S_0 represents the initial uncertainty.

Additionally, we have streamlined the text to reduce redundancy while maintaining a clear and accessible presentation of the model.

Comment 6 *I'm pleased to see they have a dataset of 2,857 state-level polls. Nice.*

Yes, we are also excited about this dataset. These polls provide a valuable foundation for our contribution but we will also make the data available and hope it can be used for future research.

Comment 7 *The authors say that "The performance of our model demonstrates that this approach works well." I would ask, "Compared to what?" e.g., Suppose a classic observational design set up with measurement level of the variables taken into account, and a regression model formulated on the specification offered (which is, at bottom, straightforward). If authors could show what they did clearly shows improvement in, say accuracy, that would give their argument more weight.*

Thank you for your comment. In response, we have now made it clearer that the Bayesian approach in our model primarily serves to estimate probability distributions, and that at its core, the model is a linear regression formulated on standard specifications.

We also explicitly compare our model's performance to several other established forecasting models. In section 5.2, we discuss how our model consistently outperforms a fundamentals-only model, particularly as election day nears. Furthermore, our model compares favorably to other election forecasting models. For instance, we find our MAE to be competitive with those reported in prior studies, such as Jennings & Wlezien (2018), Shirani-Mehr et al. (2018), and Munzert et al. (2017), demonstrating that our approach delivers strong forecasting accuracy despite the challenges in forecasting state elections.

> Our evaluation shows that the performance of the model is comparable to, and in some cases better than, other established election forecasting models. For instance, Jennings & Wlezien (2018) report an MAE of 2.7pp for presidential elections and 1.8pp for legislative elections, with performance varying based on the electoral system. In single-member district (SMD) systems, the MAE tends to be higher (2.3pp), whereas proportional representation systems have lower MAEs (1.6pp). Our model's performance is in line with these findings, particularly at the two-week and two-day lead times.
>
> Similarly, Shirani-Mehr et al. (2018) found a survey error, measured by root mean square error (RMSE), of approximately 3.5pp, about twice the size of the margins of error typically reported by polling organizations. In contrast, our model achieves much smaller errors when using a model that combines polling and fundamentals, where the RMSE improves to up to 2.06pp two days prior to the election.
>
> A direct comparison to forecasts for German federal elections shows that our model also performs well. Munzert et al. (2017) found that the RMSE for structural models in German federal elections ranged from 2.54pp to 1.98pp, depending on the proximity to election day. In the last few days before the election, models that include polling data showed substantial improvements, with the RMSE shrinking to as low as 1.69pp. Our model, similarly, improves substantially as the election approaches having an RMSE of 2.06pp two days before the election, comparable to the national election forecast.

Comment 8 *Figure 2. "The grey rhombi represent elections." I learned what a rhombi was, but the graphics do not appear to this reader to enlighten. I should add I am not against visual presentations, but these suffer.*

Thank you for pointing this out. To improve the readability of the figure, we now refer to the shape as "diamonds" and added a description of the figure in the main text. The figure is printed below.

> The grey diamonds represent election dates and the black rectangular sections indicate available polling data. The larger the rectangles, the more polling data is available for a time segment.

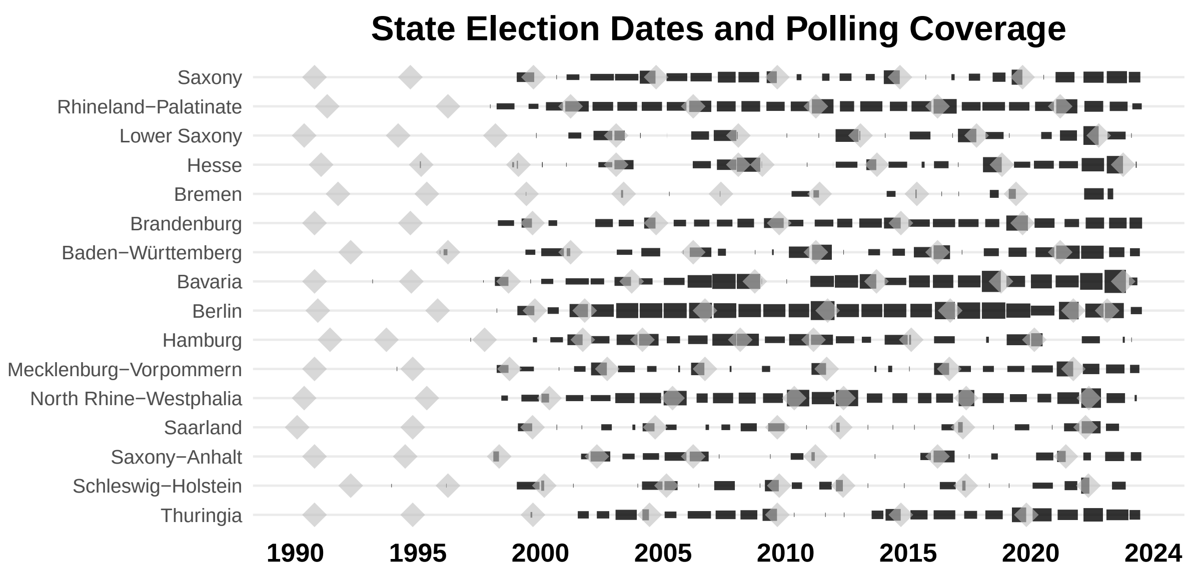

Figure 2: Poll coverage over time by state. The grey diamonds in the background represent election dates. The black segments indicate polling coverage; the larger the segments, the more polling data are available in a given year.

Comment 9 *This section deals with the "Declaration of generative AI..." I applaud the authors for making this declaration. However, I am made uneasy by the following statement: "During the preparation of this work the author(s) used ChatGPT 4o in order to copy-edit written paragraphs of text."*

Thank you for highlighting this concern; we understand the unease with AI copy editing. In response, we would first like to emphasize that our declaration follows the template provided in the journal's Guide for Authors.

Furthermore, we would like to clarify that AI was not used for text generation. Instead, it was only employed to provide suggestions for improving readability. All suggested modifications were carefully reviewed and manually edited by at least two authors to ensure accuracy and appropriateness before implementation in the manuscript.

Response to Reviewer 2

Comment 10 *First, I want to thank the editors for the opportunity to review this manuscript entitled "An Election Forecasting Model for Subnational Elections." I also wish to congratulate the authors for their fine work. This article examines the application of a Bayesian forecasting model to predict German state elections using data covering the 1990–2024 period. As mentioned by the authors, despite their significant relevance in multi-level systems, subnational elections have been relatively understudied in the forecasting literature. The paper addresses the difficulties of forecasting subnational elections, which are affected by limited polling data and distinctive regional politics. Their model relies on fundamental variables and polling information. Among other things, the authors show that a polling-only model yields more pronounced effects than the fundamentals-only model, especially as election day nears. This supports the notion that the level of party support reflected in vote intentions polls is largely driven by underlying fundamentals, capturing their influence as the election draws closer. One of the main benefits of Bayesian methods is their capacity to manage uncertainty and integrate various information sources. In regions with limited polling data, especially those with small sample sizes or high variability, Bayesian models can "borrow strength" from higher-level data, such as state or national trends, or from previous election results.*

I believe the authors submitted a methodologically sound paper with a clear empirical demonstration. I only have a two comments/suggestions for them. I hope they will find them useful and will be given the opportunity to revise their manuscript as I believe their paper would make a worthwhile contribution to the literature.

We thank the reviewer for the kind words, and valuable suggestion that we address below.

Comment 11 *I would like to see how much of the "heavy-lifting" is done by the Bayesian approach. Is there a way for the authors to compare their own approach with*

> *something like a seemingly unrelated regressions model (see, e.g., Mongrain 2019)? I think the authors' contributions would be even clearer if they were able to show that their model outperform other (simpler) approaches.*

We appreciate the reviewer's suggestion and the opportunity to clarify our approach. The Bayesian framework supports our model in two key ways: it enables the construction of a dynamic model for the poll-based component and provides a probabilistic interpretation of our forecasts. However, we do not consider the Bayesian framework to be doing the "heavy lifting," as it still relies on linear regression and employs relatively uninformative priors. To clarify its role, we have added a footnote explaining the specific advantages of this approach.

> We employ Bayesian estimation for its intuitive probabilistic interpretation of credibility intervals, which allows direct probability statements about events—unlike conventional confidence intervals. This approach is particularly effective in communicating uncertainty to a general audience. For instance, we report 5/6 credible intervals, meaning that the vote share of each party has a 5-in-6 probability of falling within this range—akin to rolling any number except a six on a fair dice. This enhances interpretability and facilitates communication, as demonstrated in recent Bayesian election forecasts (Stoetzer et al. 2019; Kang & Oh 2024; Chen et al. 2023).

Regarding alternative modeling approaches, we considered more complex regression techniques, such as a Seemingly Unrelated Regression (SUR) model or a Dirichlet regression model. However, these approaches present significant challenges in our context due to the varying number of parties across elections. In particular, the SUR model would require election-specific covariance structures (for election with the same number of parties), which is cumbersome to implement and difficult to estimate with limited data. While we have worked with Dirichlet regression models for election forecasting in other contexts, we found no clear advantages in this setting.

Given that this is the first forecasting model developed for subnational elections in Germany, we prioritized a parsimonious approach that balances methodological rigor with practical applicability. Nonetheless, we recognize the potential value of these alternative methods and now discuss their trade-offs in the manuscript, referencing relevant literature that has applied them in election forecasting.

> Looking ahead, there are several ways to extend and refine the model. Alternative modeling approaches, such as Seemingly Unrelated Regression (SUR) and Dirichlet regression, could be explored in future research. The SUR model allows for correlated errors across party forecasts, which may be useful in some election forecasting contexts (see e.g., Mongrain 2021). However, it may also require election-specific covariance structures, as different parties compete in different elections, making estimation challenging with limited data. Similarly, Dirichlet regression models are designed for compositional data and can account for the fact that vote shares sum to 100% (see e.g., Hanretty 2021; Stoetzer et al. 2019). Based on our experience with Dirichlet forecasting models, we have often found transformations of the dependent variable, such as the log-ratio approach used here, to be more practical, but future research might prove otherwise.

We appreciate the reviewer's insightful comment and hope this clarification strengthens our contribution.

Comment 12 *The model relies on a number of fundamental variables. Economic conditions (or change in economic conditions) are generally considered as one of the most important factor in structural forecasting models. This factor is absent from the authors' model. There might be good reasons for this; perhaps, state-level economic data are not available for all election years. However, if the necessary data is available, I would encourage the authors to take this factor into account. For a recent example using state-level data to forecast US presidential election outcomes, see Enns et al. (2024).*

This is an excellent point. From our perspective, the effect of economic variables in European multiparty systems is less clear than in the U.S. In multiparty systems, the responsibility for economic performance falls on the governing parties, meaning that any economic effect should be conditioned on who is in government. However, this is complicated by the presence of multiple governing parties. For this reason, established forecasting models for German federal elections do not typically include economic variables by default, as is common in U.S. models. We have explored the inclusion of various economic indicators for German federal elections in other projects but have consistently found other factors to be more predictive.

For the paper, we still have added a new specification that adds an interaction between GDP growth with government status to the fundamental model, and we continue to evaluate its impact. As expected, we find no clear effects and differences. We report the results in SM E and include a footnote in the main text.

> In this section, we present performance indicators for models that include additional variables capturing the economic conditions which might affect the electoral performance of government parties (Enns et al. 2024; Mongrain 2021). Specifically, these models additionally include growth as well as an interaction between growth and government participation. Growth is measured as the calendar and season adjusted change in real GDP. Otherwise, the models are identical to the ones presented in the main part. Table E.1 reports the average errors; compared to those from the models without economic variables, the mean errors indicate that there is little difference in performance when including these variables. Figure E.27 shows the posterior distributions which fall close to zero for the economic growth variables.

We refer to these additional evaluations in the main text.

> In SM E, we present additional models including variables to account for economic conditions at the time of an election. We include an interaction with government participation of parties as the responsibility for the economic situation might be attributed to the performance of the government (Enns et al. 2024; Mongrain 2021).

Given that the addition does not improve upon the model, and is not preregistered for the ex-ante forecasts, we decided to not integrate it in the main specification. We hope the reviewer agrees with this decision.

Response to Reviewer 3

Comment 13 *The paper explores election forecasting for subnational elections, focusing on German state elections from 1990 to 2024. It presents a Bayesian forecasting model that combines polling data and fundamental variables to address challenges unique to subnational elections, such as limited polling and localized political dynamics. This is a very good paper. I have outlined some comments* *that are not focused on the content itself but rather on the presentation. I believe the paper needs to be streamlined to enhance readability. Below are my suggestions for the author(s), which I hope will help further improve the paper.*

We thank the reviewer for this kind assessment of our paper.

Comment 14 *The introduction states that state elections are important and that research has so far concentrated on national elections (I agree with both points). However, it does not explain why it is important to present an election forecasting model for state elections. This argument is entirely missing in the introduction.*

We appreciate the reviewer's comment and acknowledge that the introduction lacked a clear explanation of why forecasting subnational elections is important. From our point of view, the value of such forecasting models lies not only in identifying local political dynamics, but also in evaluating the reliability of forecasting approaches and their methodological and theoretical foundations.

To address this, we have revised the introduction to explicitly highlight this point:

> However, while they are often analysed in the national context to identify broader political shifts, dedicated forecasting models for subnational elections remain rare. Developing such models not only provides insights into subnational political dynamics but also helps evaluate the reliability of forecasting approaches and their theoretical and methodological foundations beyond national contexts.

We believe this addition strengthens the motivation for our study and clarifies the contribution of subnational election forecasting.

Comment 15 *Regarding the introduction, and also the presentation of the model, I believe the author(s) could be more explicit about how the new model addresses the challenges of forecasting state election outcomes.*

We appreciate the opportunity to clarify how our approach addresses the challenges of forecasting state election outcomes. One challenge is the limited availability of polling data in many subnational contexts. To address this, our dynamic linear model (DLM) enables the interpolation of latent support over time, even when polling data is sparse, by leveraging the underlying structure of electoral dynamics.

Another challenge is the variability across state elections, particularly differences in party competition and electoral conditions. Our model is formulated at the party-election level, allowing it to be applied regardless of which parties compete in a given election. This structure makes it possible to identify common patterns and predictors that generalize across different contexts and parties.

We now explicitly discuss these aspects in the manuscript to better highlight the advantages of our approach.

> It addresses key challenges by incorporating a dynamic polling model that enables the interpolation of latent support before the election even in contexts with sparse polling data. Additionally, by structuring the model at the party-election level, we account for variability across elections with different party compositions while still identifying stable patterns.

Comment 16 *Regarding the introduction and the conclusion: Is the main aim of this contribution to advance the literature on state-level election forecasting, or is it to present a new model applicable to national elections (using state-level elections as a test case)?*

We appreciate the reviewer's question and the opportunity to clarify the focus of our contribution. The primary aim of this study is to advance the literature on state-level election forecasting, addressing the specific challenges of forecasting subnational elections. While the model itself can be applied to national elections, similar approaches already exist at that level. To highlight this distinction, we now elaborate on this point in greater detail in the discussion section.

> While the model could, in principle, be applied to national elections, existing approaches at that level already provide well-established forecasting methods. Our contribution lies in adapting and refining these techniques for subnational elections, where forecasting models remain underdeveloped, and our results demonstrate that the theoretical and methodological foundations perform well in this context.

Comment 17 *Minor point: The introduction refers to "so-called fundamental variables." I suggest providing an explanation or definition when this term is first introduced.*

We appreciate the reviewer's suggestion and have clarified the term "fundamental variables" in the introduction. Specifically, we now provide a brief explanation of what these variables encompass and have removed the phrase "so-called" for clarity. The revised sentence reads:

> Forecasting subnational elections presents unique challenges for traditional election forecasting models, which typically depend on pre-election polls or fundamental variables—such as economic indicators, incumbency status, and government approval ratings—to predict electoral outcomes (Lewis-Beck & Dassonneville 2015b; Nadeau et al. 2020).

Comment 18 *I recommend dedicating a full section to the literature review, which is currently only a short subsection in the introduction (1.1. Existing Forecasting Models).*

Thanks for pointing this out. We moved the literature discussion in its own section. Additionally, we have extended this section by including references to additional studies on forecasting subnational elections.

> Several studies have developed structural models for forecasting subnational elections in the U.S. For example, Bardwell & Lewis-Beck (2004) presents a state-level model for forecasting U.S. Senate elections, while Linzer (2013) applies a Bayesian approach to predict U.S. presidential election outcomes at the state level. Similarly, Klarner (2010, 2018) introduce models for forecasting U.S. state legislative elections in 2010 and 2018, respectively. In the context of gubernatorial races, Hummel & Rothschild (2014) develops a model accounting for state-level fundamentals.

Comment 19 *I must admit that I found it difficult to follow the presentation of the forecasting model in section 2. I suggest that the author(s) focus on "walking the reader through" their model more clearly. For instance, while I learn that the observed fundamentals are "in a matrix X that has N rows," I am not told what the fundamentals actually are.*

We appreciate the reviewer's feedback and have thoroughly revised the section to improve clarity.

We now introduce each paragraph with a non-technical sentence. Additionally, we provide examples of fundamental variables commonly used in election forecasting to clarify the types of predictors that can be included. We have also restructured the explanation to provide detail and explanation on the modelling framework, with a particular focus on clarifying the dynamic model. Furthermore, we offer a more detailed explanation of the predictive distribution.

Please see changes in Section 3.

We hope these revisions enhance the readability of Section 3 and make the forecasting model more accessible.

Comment 20 *This is, of course, a matter of style, but I found it puzzling that the case selection (German states) is explained in section 3, and then subsections 3.1 and 3.2 present the main analyses. I recommend dedicating a new section (section 4) to the election forecasting model, covering varying lead times, evaluation based on past elections, and the ex-ante forecast.*

We appreciate this suggestion and agree that restructuring the sections will improve readability and coherence. To address this, we introduced a new section structure:

- **Section 3: The Case of German State Elections** – This section discusses German state elections as the focus of the study.
- **Section 4: Application to German State Elections** – This section presents the election forecasting model, including varying lead times, evaluation based on past elections, and the ex-ante forecast.

This adjustment ensures a more logical progression from case selection to model application, making it easier for readers to follow the analysis.

Comment 21 *I appreciate the graphs in this paper; they provide an excellent illustration of the results. Regarding Figures 4 and 6, perhaps the x- and y-axes could be adjusted so that the x-axis represents time moving closer to Election Day.*

Thank you for the feedback! We have adjusted the x-axes of Figures 4 and 6 so that they now represent time moving closer to Election Day. The updated figures are printed below and are now easier to read.

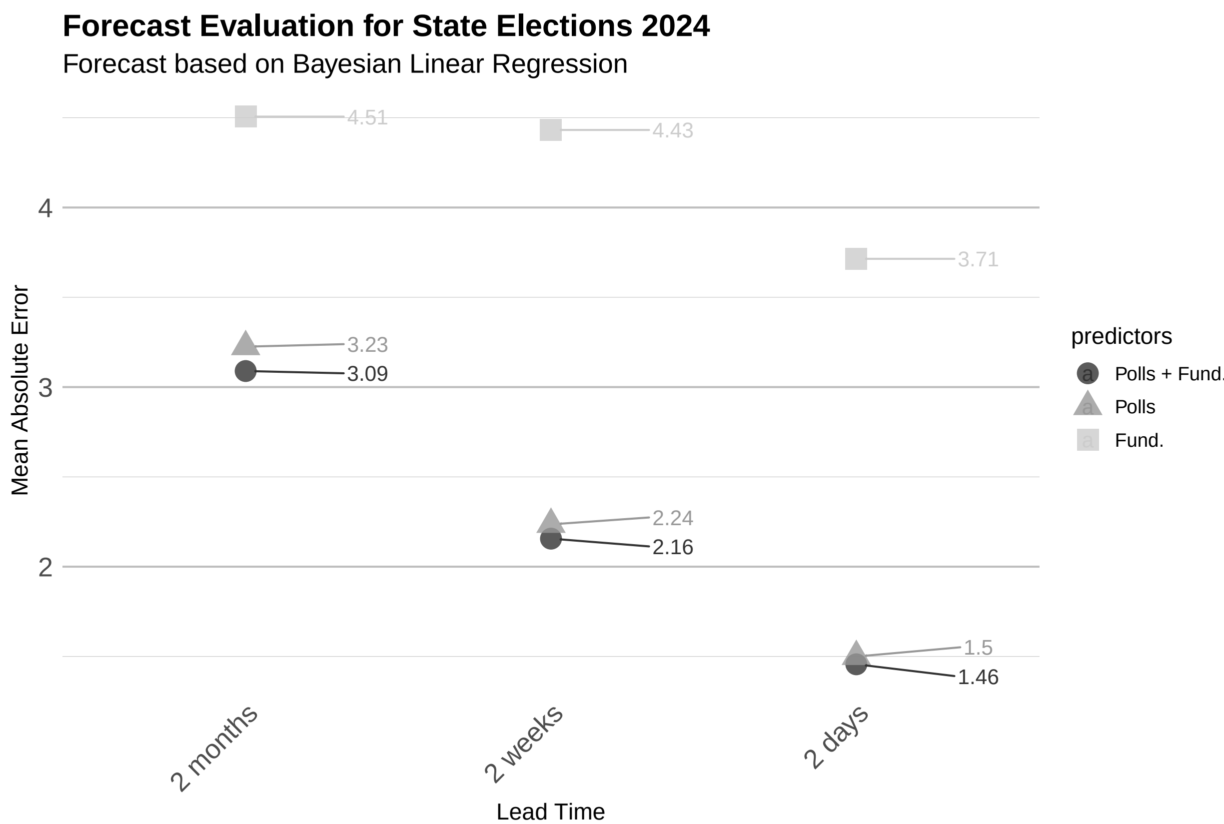

Figure 4: Mean Absolute Error (MAE) for the state elections from 2015 to 2023, comparing different model specifications. The MAE values are calculated for forecasting models using different lead times (days, weeks, and months) prior to the election. This figure demonstrates the accuracy of the model's predictions over time, with lower MAE values indicating better model performance.

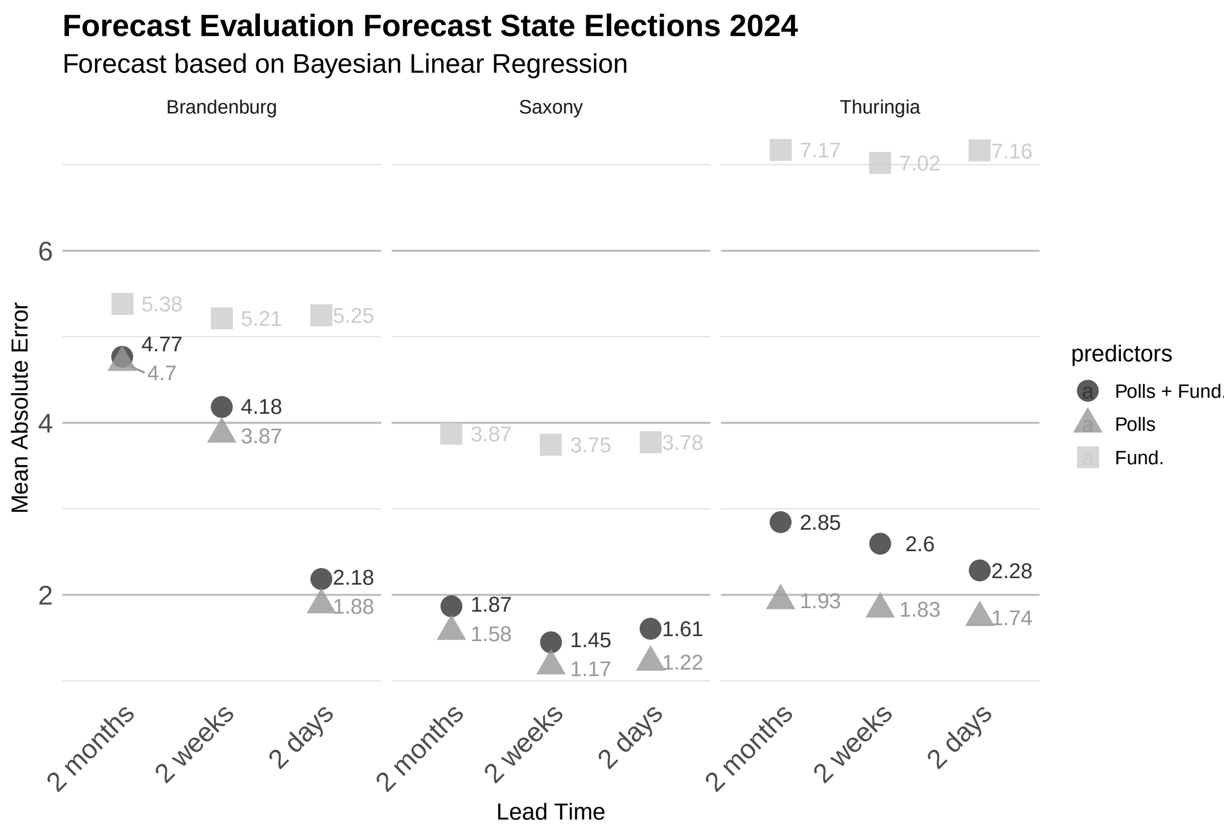

Figure 6: Mean Absolute Error (MAE) for the ex-ante forecasts of the 2024 state elections, comparing different model specifications. The MAE values are calculated for forecasting models at different time points (days, weeks, and months) before the election. This figure highlights the improving accuracy of the forecasts as the election approaches, with lower MAE values reflecting better model performance.

Comment 22 *Finally, in the conclusion, I think the paper should more explicitly state its main contribution (which I interpret as advancing state election forecasting models) and be more specific about its generalizability to other countries. Currently, the conclusion is very general, stating, "in other countries with comparable political structures." Which countries are comparable? Was Germany a typical case? Does this depend on the state structure, or does it relate to the number of polls conducted before a state election?*

We thank the reviewer for this suggestion and consequently revised the conclusion to explicitly highlight the main contribution of our study—advancing state election forecasting models.

> Our contribution lies in adapting and refining these techniques for subnational elections, where forecasting models remain underdeveloped, and our results demonstrate that the theoretical and methodological foundations perform well in this context.

We have also expanded the discussion on the generalizability of our approach. Specifically, our model is applicable to countries with legislative state elections, where both structural predictors and polling data play a role in forecasting outcomes. Examples of comparable such countries include Canada (Provincial and territorial elections), Switzerland (Cantonal elections), and Spain (Regional elections in autonomous communities), where regional elections occur regularly and involve multiple competing parties. While Germany serves as a strong test case due to its combination of frequent state elections and availability of polling data, our framework can also be adapted to contexts with fewer pre-election polls by placing greater weight on fundamental predictors. We now explicitly discuss these aspects in the conclusion.

> The potential for applying this model to other subnational election contexts, outside of Germany, is promising. The consistent performance across different German states suggests that similar models could be adapted for federal or regional elections in other countries with comparable political structures. Many federal democratic systems feature subnational elections that shape regional governance and national politics. Comparable cases include Canada (provincial elections), Switzerland (cantonal elections), and Spain (autonomous com-

munity elections), where multi-party competition and varying polling availability present similar forecasting challenges.

Bibliography

Allers, M., Rienks, H., & de Natris, J. (2022). When are parties punished for serving in a coalition government? *Electoral Studies*, *79*, 102516.

Bardwell, K., & Lewis-Beck, M. S. (2004). State-level forecasts of us senate elections. *PS: Political Science & Politics*, *37*(4), 821–826.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32.

Chen, Y., Garnett, R., & Montgomery, J. M. (2023). Polls, context, and time: a dynamic hierarchical bayesian forecasting model for us senate elections. *Political Analysis*, *31*(1), 113–133.

Eggers, A. C., & Spirling, A. (2017). Incumbency Effects and the Strength of Party Preferences: Evidence from Multiparty Elections in the United Kingdom. *The Journal of Politics*, *79*(3), 903–920.
URL `https://doi.org/10.1086/690617`

Enns, P. K., Colner, J., Kumar, A., & Lagodny, J. (2024). Understanding biden's exit and the 2024 election: The state presidential approval/state economy model. *PS: Political Science & Politics*, (pp. 1–8).

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2020). *rstanarm: Bayesian applied regression modeling via Stan*. R package version 2.21.1.
URL `https://mc-stan.org/rstanarm/`

Hanretty, C. (2021). Forecasting multiparty by-elections using Dirichlet regression. *International Journal of Forecasting*, *37*(4), 1666–1676.

Hummel, P., & Rothschild, D. (2014). Fundamental models for forecasting elections at the state level. *Electoral Studies*, *35*, 123–139.

Jennings, W., & Wlezien, C. (2018). Election polling errors across time and space. *Nature Human Behaviour*, *2*(4), 276–283.

Kang, S., & Oh, H.-S. (2024). Forecasting south korea's presidential election via multiparty dynamic bayesian modeling. *International Journal of Forecasting*, *40*(1), 124–141.

Klarner, C. (2010). Forecasting the 2010 state legislative elections. *PS: Political Science & Politics*, *43*(4), 643–648.

Klarner, C. E. (2018). 2018 state legislative election forecasts. *PS: Political Science & Politics*, *51*(S1), 21–27.

Lewis-Beck, M. S., & Dassonneville, R. (2015a). Comparative election forecasting: Further insights from synthetic models. *Electoral Studies*, *39*, 275–283.

Lewis-Beck, M. S., & Dassonneville, R. (2015b). Forecasting elections in Europe: Synthetic models. *Research & Politics*, *2*(1), 2053168014565128.

Lewis-Beck, M. S., Nadeau, R., & Bélanger, (2016). The british general election: Synthetic forecasts. *Electoral Studies*, *41*, 264–268.

Linzer, D. A. (2013). Dynamic bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, *108*(501), 124–134.

Mongrain, P. (2021). 10 downing street: who's next? seemingly unrelated regressions to forecast uk election results. *Journal of Elections, Public Opinion and Parties*, *31*(1), 22–32.

Munzert, S., Stötzer, L., Gschwend, T., Neunhoeffer, M., & Sternberg, S. (2017). Zweitstimme.org. Ein strukturell-dynamisches Vorhersagemodell für Bundestagswahlen. *Politische Vierteljahresschrift*, *58*(3), 418–441.
URL `http://www.jstor.org/stable/26427772`

Nadeau, R., Dassonneville, R., Lewis-Beck, M. S., & Mongrain, P. (2020). Are election results more unpredictable? A forecasting test. *Political Science Research and Methods*, *8*(4), 764–771.

Petris, G. (2010). An R package for dynamic linear models. *Journal of Statistical Software*, *36*(12), 1–16.
URL `http://www.jstatsoft.org/v36/i12/`

Shirani-Mehr, H., Rothschild, D., Goel, S., & Gelman, A. (2018). Disentangling bias and variance in election polls. *Journal of the American Statistical Association*, *113*(522), 607–614.

Stoetzer, L. F., Neunhoeffer, M., Gschwend, T., Munzert, S., & Sternberg, S. (2019). Forecasting Elections in Multiparty Systems: A Bayesian Approach Combining Polls and Fundamentals. *Political Analysis*, *27*(2), 255–262.

Stoetzer, L. F., & Orlowski, M. (2020). Estimating coalition majorities during political campaigns based on pre-election polls. *Journal of Elections, Public Opinion and Parties*, *30*(1), 126–137.

Walther, D. (2015). Picking the winner (s): Forecasting elections in multiparty systems. *Electoral Studies*, *40*, 1–13.

West, M., & Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer, 2nd ed.